



Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

Downlink scheduling in CDMA data networks

A. Bedekar, S.C. Borst, K. Ramanan, P.A. Whiting, E.M. Yeh

Probability, Networks and Algorithms (PNA)

PNA-R9910 October 31, 1999

Report PNA-R9910
ISSN 1386-3711

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

CWI is the National Research Institute for Mathematics and Computer Science. CWI is part of the Stichting Mathematisch Centrum (SMC), the Dutch foundation for promotion of mathematics and computer science and their applications.

SMC is sponsored by the Netherlands Organization for Scientific Research (NWO). CWI is a member of ERCIM, the European Research Consortium for Informatics and Mathematics.

Copyright © Stichting Mathematisch Centrum
P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

Downlink Scheduling in CDMA Data Networks

A. Bedekar

University of Washington, Seattle WA, USA

S.C. Borst*

CWI

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

K. Ramanan, P.A. Whiting

Bell Laboratories, Lucent Technologies

Murray Hill, NJ 07974, USA

E.M. Yeh

Massachusetts Institute of Technology, Boston MA, USA

ABSTRACT

We identify optimality properties for scheduling downlink transmissions to data users in CDMA networks. For arbitrary-topology networks, we show that under certain idealizing assumptions it is optimal for a base station to transmit to only one data user at a time. Moreover, for data-only networks, we prove that a base station, when on, should transmit at maximum power for optimality. We use these two properties to obtain a mathematical programming formulation for determining the optimal transmission schedule in linear data-only networks, with time allocations playing the role of decision variables. The optimality conditions imply that there exist (i) subsets of outer users on either side of the cell that should be served when only the neighboring base station on the opposite side is on; (ii) a subset of inner users in the center of the cell that should be served when both neighbors are on; (iii) a subset of users in the intermediate regions that should receive transmissions when both neighbors are off. Exploiting these structural properties, we derive a simple search algorithm for finding the optimal transmission schedule in symmetric scenarios. Numerical experiments illustrate that scheduling achieves significant capacity gains over conventional CDMA.

1991 Mathematics Subject Classification: 60K25, 68M20, 90B12, 90B22.

Keywords and Phrases: achievable throughput, CDMA data networks, downlink scheduling, inter-cell scheduling, intra-cell scheduling, throughput maximization.

Note: Work of the second author carried out in part under the project PNA2.1 “Communication and Computer Networks”.

*Also affiliated with Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands, and Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, USA

1 Introduction

The continual growth of wireless communications, combined with the popularity of Internet applications such as electronic mail, file transfers, and Web browsing promise great potential for wireless high-speed data services. This has fueled a strong demand for advanced wireless systems that offer high data rates and support a wide range of services.

Existing wireless systems, which were originally expected to provide only voice services, offer rather limited capabilities to support data applications, see Alanko *et al.* [2] and Mathias [9]. These limitations reflect fundamental differences between voice and data communications. The traffic characteristics for example are notably different, as data traffic is typically more bursty than voice, with transmission bursts alternating with long silence periods. The quality-of-service requirements are also radically different, in that data traffic is usually less delay sensitive, but may have higher throughput requirements than voice during burst periods. These differences suggest that it may be beneficial to control the transmission rate for data users in a highly adaptive manner, rather than provide a constant guaranteed rate as for voice calls. This induces a radical departure from traditional resource management strategies which were specifically designed to support voice connections.

The majority of present TDMA systems for example use fixed channel assignment to provide a dedicated connection for each voice call. The bursty characteristics of data traffic seem however to require dynamic channel assignment strategies operating on the time scale of burst periods in order to achieve high spectrum efficiency, see for instance Andrews *et al.* [3] and Cimini *et al.* [4]. In current CDMA systems, which support a large number of low-rate voice calls, power control algorithms crucially rely on the fact that interference varies only moderately due to statistical averaging. With a small number of high-rate data connections, however, interference may fluctuate dramatically unless carefully controlled.

In view of the difficulties mentioned above, changes to existing CDMA standards have been proposed so as to permit high-rate data traffic. One natural approach is to schedule the transmission of data users' signals so as to avoid interference. The main idea is that the increased transmission rate at lower interference more than compensates for the loss in available transmission time. This is the approach advocated in Pottie [10]. The simplest scheme in this direction is *intra-cell scheduling*, where data users' packets are transmitted one at a time within each cell. There is no attempt to coordinate these transmissions between base stations. An approach along these lines was examined in Ramakrishna [12] and Holtzman & Ramakrishna [5] for the uplink in a mixed voice and data system.

An alternative, more complex approach is *inter-cell scheduling*, where base stations coordinate their transmissions to data users. A version of this approach has been examined and standardized in revision B of the IS-95 standard, see for instance I & Nanda [6] and Knisely *et al.* [7]. On the uplink, this coordination involves the reporting of pilot strength information

back to the base station by each mobile when requesting a data burst. These measurements, along with the loadings of cells in the neighborhood of the mobile, are used to determine whether the burst will be admitted. A similar procedure is performed on the downlink. Admission is granted at a rate nominated by the base station, see [6]. Performance analysis is carried out primarily through simulations.

In the present paper we use analytic techniques to identify optimal transmission schedules for the downlink in CDMA networks. The downlink is particularly relevant to consider because it is commonly expected that data traffic will be asymmetric with the bulk of the traffic directed from the base station to the user. In contrast to the uplink, downlink transmissions can be synchronized, resulting in reduced interference between users within the same cell. This makes the case for scheduling on the downlink less clear, and motivates the need for a careful analysis.

We adopt a standard model for the transmission rate received by users as a function of the power level at the base stations, see Viterbi [13]. Under some idealizing assumptions, we show that it is optimal for a base station to transmit to only one data user at a time. Moreover, for data-only networks we prove that a base station, when on, should transmit at maximum power for optimality. These two properties imply that the interference experienced by a user is determined by the on/off state of the surrounding base stations. Hence the throughput may be expressed as a function of the fractions of time that a user receives transmission when neighboring base stations are on or off. This allows the optimal transmission schedule in linear data-only networks to be formulated as a mathematical program, with time allocations playing the role of decision variables.

As noted above, our results rely on certain idealizing assumptions that facilitate the analysis. We ignore for example any constraint placed on the maximum transmission rate by the available bandwidth because of a minimum spreading factor requirement. In most of the analysis, we also neglect the self-noise of the Rake (or matched filter) receiver in the presence of self-noise, see Proakis [11]. A model that takes into account these constraints would yield lower actual transmission rates. In that sense, our results provide bounds on the achievable performance gains from scheduling in less ideal circumstances. More importantly, we expect that some of the fundamental insights from the analysis extend beyond the formal scope of the model. In particular, we believe that the qualitative properties of the optimal transmission schedules reflect desirable features of candidate scheduling algorithms for more complex practical systems which may not lend themselves to exact analysis. This is borne out by our numerical results, which show significant capacity gains over conventional CDMA even when some practical constraints are taken into account.

The remainder of the paper is organized as follows. In Section 2 we introduce the basic model. In Section 3 we identify two crucial optimality principles for scheduling downlink transmissions to data users in networks of arbitrary topology. These properties are used in Section 4

to obtain a mathematical programming formulation for determining the optimal transmission schedule in linear data-only networks. We then give a detailed characterization of the structure of the optimal schedule. Exploiting these structural properties, we derive in Section 5 a simple search algorithm for finding the optimal schedule in symmetric scenarios. In Section 6 we describe some numerical experiments which were performed to evaluate the performance gains from intra-cell and inter-cell scheduling. We conclude with some final remarks in Section 7.

2 Model description

In this section we describe our basic model. We adopt a standard model for the transmission rate received by users as a function of the power level at the base stations, see Viterbi [13, Chapter 6]. Let W be the system bandwidth and R the transmission rate required by a user. If P is the power received by the user from its own base station, and I the total noise power or interference experienced by that user, then the energy-per-bit to noise density ratio, E_b/N_0 , is given by

$$\frac{E_b}{N_0} = \frac{WP}{RI}. \quad (1)$$

In order for the user to be able to decode the base station's signal with an acceptable probability of error ε , it is necessary that $E_b/N_0 \geq \gamma$, where $\gamma = \gamma(\varepsilon)$ is some threshold which is determined by the probability of error ε . Substituting (1) in this last inequality, we find that for a given received power P , the achievable rate is

$$R = \frac{WP}{\gamma I}. \quad (2)$$

The relation (2) shows that the achievable rate depends on the power received by the user from its own base station, as well as the total interfering noise power experienced by that user due to background noise and transmissions to other users in the network.

We now introduce some notation and describe the specifics of the model. We consider a network of base stations with arbitrary topology. We focus on data users, although in some of the analysis we allow for the presence of voice users as well. Assume that the base stations are numbered, and let P^k be the total power at which the k^{th} base station is transmitting at some fixed time t . Denote by \mathcal{G}_i^{km} the path loss from base station m to data user i in cell k . Then the total interference experienced by data user i in cell k due to transmissions of other base stations is

$$I_i^k = \sum_{m \neq k} P^m \mathcal{G}_i^{km}.$$

Let P_V^k and P_D^k be the total power at which base station k is transmitting to voice and data users respectively, so that $P^k = P_V^k + P_D^k$. Suppose there are M^k data users in the k^{th} cell. Define ϕ_i^k as the fraction of data power P_D^k that is transmitted by base station k to data user i ,

so that $\sum_{i=1}^{M^k} \phi_i^k = 1$. Then the total interference caused to data user i by its own base station's transmissions is

$$f_i^k [h_i^k \phi_i^k P_D^k + (1 - \phi_i^k) P_D^k + P_V^k] \mathcal{G}_i^{kk},$$

where f_i^k is the orthogonality factor and h_i^k a self-noise coefficient. We assume $h_i^k \leq 1$, which amounts to saying that a user experiences no more interference from its own signal than from the signals of other users in the cell. Plugging the above expressions in (2), we find that the rate received by data user i in cell k at time t is

$$R_i^k = \frac{W}{\gamma} \frac{\phi_i^k P_D^k \mathcal{G}_i^{kk}}{I_i^k + f_i^k [h_i^k \phi_i^k P_D^k + (1 - \phi_i^k) P_D^k + P_V^k] \mathcal{G}_i^{kk} + \eta}, \quad (3)$$

where η is the background noise.

The orthogonality factor f_i^k represents the fraction of power transmitted by base station k that appears as interference to user i . If the base station uses orthogonal codes to transmit to distinct users, then intra-cell interference is virtually eliminated when the channel is Gaussian, which corresponds to $f_i^k = 0$. When there is multi-path, however, this form of interference is only partially reduced, which implies $f_i^k \in (0, 1)$. The orthogonality factor has a significant impact on performance, as discussed in greater detail in Sections 3.1 and 7.

We assume that data users have throughput demands (as opposed to instantaneous rate requirements) defined in terms of average rate requirements over time periods of a certain fixed duration, e.g. a number of frames. This duration must be short compared to the maximum delay tolerance of the data users. It should also be short with respect to the time scale of queue length dynamics and changes in the user population, so that throughput demands can be assumed constant during each period. The user positions and throughput demands may change from one period to the next due to mobility, admission of calls, termination of calls, and queues draining or building up. On the other hand, the time period must be large enough to allow for the communication and computation involved in scheduling the data transmissions. Thus there is a fundamental trade-off between the potential throughput of a scheme and its latency measured in terms of the length of a scheduling period. In our analysis we consider one such time period, which we set without loss of generality to be of unit length.

A vital issue that arises in dynamic scenarios is that of admission control, i.e., which combinations of users can be admitted while satisfying quality-of-service requirements. A related issue is that of rate allocation, or the determination of how throughput targets should be set in response to queue length dynamics so as to meet the quality-of-service requirements. In order to answer these questions, it is crucial to know what rate vectors can be achieved and how best to schedule transmissions so as to achieve those rates. These are the two problems addressed in the remainder of the paper.

We close the section with a brief discussion of some of the assumptions implicit in our model. These modeling assumptions permit a transparent analysis of optimal scheduling strategies, but ignore certain practical constraints. We ignore for example any constraint placed on the maximum transmission rate by the available bandwidth due to a minimum spreading factor requirement. In practice, the requirement to spread the signal limits the peak transmission rate to a small fraction of the available bandwidth. In addition we do not consider any synchronization or pilot signals. It is expected that any wideband system would provide synchronization signals in order to avoid repeated acquisition of the spread spectrum signal prior to transmission. In most of the analysis, we also neglect the self-noise of the Rake (or matched filter) receiver in the presence of multi-path, see Proakis [11]. This self-noise ultimately determines the maximum information rate of the Rake which is thus interference limited. However, self-noise can be reduced, at least if the Rayleigh fading frequency is low, by using alternate receiver designs. One such design uses a spread sequence matched filter in combination with a decision feedback equalizer to reduce self-noise arising from multi-path, see Abdulrahman *et al.* [1]. The minimum mean square error receiver, on the other hand, uses a chip matched filter instead, see Madhow & Honig [8].

3 Optimality properties

We now derive two crucial optimality principles for scheduling downlink transmissions to data users. We use these properties in the next section to obtain a mathematical programming formulation for determining the optimal transmission schedule in linear data-only networks.

3.1 Intra-cell scheduling

We first show that it is better for a base station to transmit to only one data user at a time rather than transmit to several such users simultaneously.

Consider a time interval during which the total power level P^k at each base station remains constant, and focus on an arbitrary base station k^* in the network. For notational convenience we drop the index k^* for the remainder of the section.

Suppose that throughout the interval a proportion ϕ_i of the total data power P_D is allocated to the i^{th} data user so that $\sum_i \phi_i = 1$. From (3), the rate received by user i is

$$R_i^A = \frac{W}{\gamma} \frac{\phi_i P_D \mathcal{G}_i}{I_i + f_i [h_i \phi_i P_D + (1 - \phi_i) P_D + P_V] \mathcal{G}_i + \eta}.$$

The throughput per unit time is also R_i^A since user i receives this rate throughout the interval.

Now suppose instead that each user i is allocated the total data power P_D , but only for a fraction ϕ_i of the interval. In other words, ϕ_i is used as a time-sharing rather than a power-sharing parameter. Note that then the total data power used throughout the interval remains

the same (equal to P_D), and hence voice users within the cell experience the same interference as before. The same is true of all users, both voice and data, in other cells. So none of the users other than the data users in the cell under consideration are affected.

During the period that the base station transmits to user i , no power is allocated to other data users in the cell. Thus user i receives no interference from other data users within the cell, and so the rate received during that period is

$$R_i^B = \frac{W}{\gamma} \frac{P_D \mathcal{G}_i}{I_i + f_i [h_i P_D + P_V] \mathcal{G}_i + \eta}.$$

Since user i receives this rate for only a fraction ϕ_i of the time, and receives nothing otherwise, the throughput per unit time is $\phi_i R_i^B$.

Comparing the two schemes, we find that the throughput ratio is

$$\frac{R_i^A}{\phi_i R_i^B} = 1 + \frac{(1 - h_i)(1 - \phi_i) f_i P_D \mathcal{G}_i}{I_i + f_i [h_i P_D + P_V] \mathcal{G}_i + \eta} \geq 1.$$

If there is only one data user in the cell, (so that $\phi_i = 1$), or there is perfect orthogonality and no multipath ($f_i = 0$), then the above ratio is equal to 1. In that case, there is clearly no distinction between the two schemes. Otherwise, the above ratio is strictly greater than 1, assuming $h_i < 1$.

Thus we conclude that for data users, it is better to transmit to one user at a time rather than transmit to several users simultaneously. This conclusion holds as long as $h_i < 1$, regardless of the topology of the network, the presence of voice users in the cell, and user locations.

3.2 Maximum power setting

In the previous subsection we showed that it is optimal for a base station to transmit to only one data user at a time. We now determine at what power level a base station should transmit for optimality. We focus on a data-only scenario (which corresponds to $P_V^k = 0$ and $P_D^k = P^k$ for all k) and neglect self-noise (so that $h_i^k = 0$).

Once again consider a time interval during which the power level P^k used at each of the base stations remains constant. By the result of the previous subsection, we may assume that when base station k is on, the total power P^k is allocated to only one user. Suppose the identity of that user does not change during the interval under consideration. Then by (3) the rate received by that user is

$$R_i^k = \frac{W}{\gamma} \frac{P^k \mathcal{G}_i^{kk}}{I_i^k + \eta}, \quad (4)$$

where the external interference I_i^k is given by

$$I_i^k = \sum_{m \neq k} P^m \mathcal{G}_i^{km}.$$

Now consider an arbitrary base station k^* , and suppose that instead of transmitting at a constant power level P^{k^*} , it uses full power $P_{\max}^{k^*}$ during a fraction $P^{k^*}/P_{\max}^{k^*}$ of the interval, and zero power for the remainder of the time. Then user i in cell k^* receives the same *average power*, and experiences the same interference. Consequently, since rate is linear in the received power, the average rate received by user i is not altered. Users in other cells receive the same power, while experiencing the same *average interference*. Hence the new average rate received by other users is no less than the old one since the rate R_i^k is convex in the total interference I_i^k in (4).

This shows that in the absence of self-noise and voice users, it is optimal for a base station, when on, to transmit at maximum power. This observation relies on the fact that the rate R_i^k is linear in the power level P^k and convex in the interference I_i^k , and holds irrespective of the particular form of that relationship.

In the presence of self-noise or voice users, by (3) the total interference experienced by user i in cell k is equal to

$$I_i^k + f_i^k [h_i^k P_D^k + P_V^k] \mathcal{G}_i^{kk} + \eta.$$

An increase in data power transmitted to user i would cause increased interference to itself (due to the term $f_i^k h_i^k P_D^k$) as well as to the voice users. In response to this increased interference, the voice users would have to step up their power P_V^k so as to maintain their signal-to-interference ratio at the target level. Thus an increase in data power transmitted to a user triggers an increase in the interference to that user, destroying the linearity property mentioned above. For the same reason, the presence of voice users also destroys the linear dependence of the interference experienced by users in one cell on the transmit powers of users in other cells, thus allowing for the possibility of a concave dependence of rate on the transmit power of other cells. These issues are the subject of current research.

4 Linear data networks

In the previous section it was shown that for data-only networks it is optimal for a base station to transmit to only one data user at a time, and then at maximum power. What remains to be determined is when the base station should be on, and to what user it should transmit at that time. In this section we derive a mathematical programming formulation for determining the optimal transmission schedule in a linear data-only network.

Consider a linear array of cells indexed $1, \dots, K$ (from left to right). We focus on a data-only scenario, so that $P_V^k = 0$, and $P_D^k = P^k$ for all $k = 1, \dots, K$. We assume that the users within cell k are numbered $1, \dots, M_k$ (also from left to right), and that they only experience external interference from the two adjacent base stations $k-1$ and $k+1$, so that $\mathcal{G}_i^{k,l} = 0$ if $|k-l| > 1$. (To ensure that the neighbors for the two edge base stations are well-defined, we

introduce two dummy base stations 0 and $K + 1$.)

The optimality property of Subsection 3.2 implies that each base station alternates between on periods, when it transmits at maximum power, and off periods. Thus at any given time, $P^k = \delta_k P_{\max}^k$, where the variable $\delta_k \in \{0, 1\}$ indicates the on/off state of base station k .

The result of Subsection 3.1 implies that when a base station is on, the total power is allocated to only one user. Thus, from (3), when base station k is transmitting to user i , the received rate is

$$R_i^k = \frac{W}{\gamma} \frac{P_{\max}^k \mathcal{G}_i^{kk}}{\delta_{k-1} P_{\max}^{k-1} \mathcal{G}_i^{k,k-1} + \delta_{k+1} P_{\max}^{k+1} \mathcal{G}_i^{k,k+1} + \eta}. \quad (5)$$

Notice that the received rate only depends on the on/off configuration of the two adjacent base stations. For base station k , let us refer to the time during which only its *left* neighbor $k - 1$ is on as its L-time. Similarly, the time during which only its *right* neighbor $k + 1$ is transmitting, is called its R-time. The time during which *both* neighbors are on is referred to as its 2-time. Finally, the time during which *none* of its neighbors is transmitting is termed its 0-time. Figure 1 illustrates these scenarios for a particular cell E_0 and its two neighbors E_{-1} and E_1 .

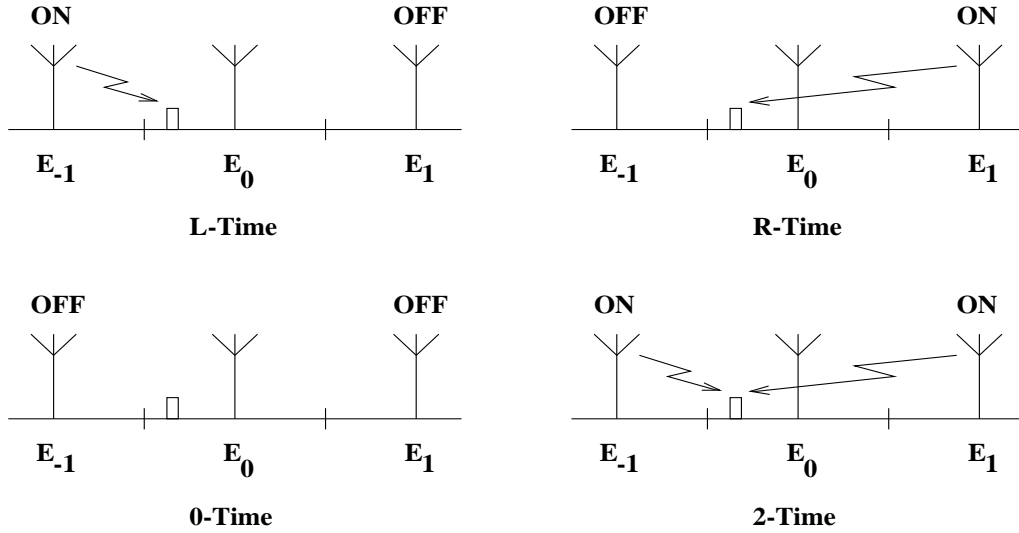


Figure 1: On/off states of cell E_0 's neighboring base stations during its L-time, R-time, 0-time, and 2-time.

Denote by τ_{ik}^0 the fraction of the unit interval during which base station k transmits to user i in 0-time. Define τ_{ik}^L , τ_{ik}^R , and τ_{ik}^2 analogously. Let $\tau_k^{0,on}$ denote the fraction of time that base station k is on in 0-time, and let $\tau_k^{0,off}$ be the fraction of time that base station k is off in 0-time. Define $\tau_k^{L,on}$, $\tau_k^{R,on}$, $\tau_k^{2,on}$, $\tau_k^{L,off}$, $\tau_k^{R,off}$, and $\tau_k^{2,off}$ similarly. By definition

$$\tau_k^{0,on} = \sum_{i=1}^{M_k} \tau_{ik}^0.$$

Similar identity relations hold for $\tau_k^{L,on}$, $\tau_k^{R,on}$, and $\tau_k^{2,on}$.

Let $P_{ik}^L \doteq P_{\max}^{k-1} \mathcal{G}_i^{k,k-1}$ be the interference that user i receives when base station $k-1$ is on. Similarly let $P_{ik}^R \doteq P_{\max}^{k+1} \mathcal{G}_i^{k,k+1}$. Then from (5) the throughput that user i receives is

$$\begin{aligned} R_i^k &= \frac{W}{\gamma} \left[\tau_{ik}^0 \frac{P_{\max}^k \mathcal{G}_i^{kk}}{\eta} + \tau_{ik}^L \frac{P_{\max}^k \mathcal{G}_i^{kk}}{P_{ik}^L + \eta} + \tau_{ik}^R \frac{P_{\max}^k \mathcal{G}_i^{kk}}{P_{ik}^R + \eta} + \tau_{ik}^2 \frac{P_{\max}^k \mathcal{G}_i^{kk}}{P_{ik}^L + P_{ik}^R + \eta} \right] \\ &= \frac{W}{\gamma} \frac{P_{\max}^k \mathcal{G}_i^{kk}}{\eta} \left[\tau_{ik}^0 + \frac{\tau_{ik}^L}{1 + P_{ik}^L/\eta} + \frac{\tau_{ik}^R}{1 + P_{ik}^R/\eta} + \frac{\tau_{ik}^2}{1 + P_{ik}^L/\eta + P_{ik}^R/\eta} \right]. \end{aligned}$$

Denoting $\beta_{ik}^L \doteq P_{ik}^L/\eta$, $\beta_{ik}^R \doteq P_{ik}^R/\eta$, and defining the normalized throughput to be

$$\tilde{R}_i^k \doteq \frac{\gamma}{W} \frac{\eta}{P_{\max}^k \mathcal{G}_i^{kk}} R_i^k,$$

we obtain

$$\tilde{R}_i^k = \tau_{ik}^0 + \frac{\tau_{ik}^L}{1 + \beta_{ik}^L} + \frac{\tau_{ik}^R}{1 + \beta_{ik}^R} + \frac{\tau_{ik}^2}{1 + \beta_{ik}^L + \beta_{ik}^R}. \quad (6)$$

Relation (6) describes the (normalized) throughput of each user as a function of the time allocation vector $\tau \doteq (\tau_{ik}^0, \tau_{ik}^L, \tau_{ik}^R, \tau_{ik}^2, i = 1, \dots, M^k, k = 1, \dots, K)$. It is clear that not all vectors τ correspond to feasible time allocations. The components of τ have to satisfy additional compatibility conditions. For instance, time conservation dictates that for all $k = 1, \dots, K$,

$$\tau_k^{0,on} + \tau_k^{0,off} + \tau_k^{L,on} + \tau_k^{L,off} + \tau_k^{R,on} + \tau_k^{R,off} + \tau_k^{2,on} + \tau_k^{2,off} = 1. \quad (7)$$

In addition, base stations k and $k+1$ are off exactly when base station k is off and has 0-time or L-time and base station $k+1$ is off and has 0-time or R-time. This yields, for $k = 1, \dots, K-1$,

$$\tau_k^{0,off} + \tau_k^{L,off} = \tau_{k+1}^{0,off} + \tau_{k+1}^{R,off}. \quad (8)$$

Similarly the time during which base station k is off and has R-time or 2-time, and base station $k+1$ is on and has 0-time or R-time coincides with the time k is off while $k+1$ is on. Consequently, for $k = 1, \dots, K-1$,

$$\tau_k^{R,off} + \tau_k^{2,off} = \tau_{k+1}^{0,on} + \tau_{k+1}^{R,on}. \quad (9)$$

In an analogous fashion, base station k is on and has 0-time or L-time and base station $k+1$ is off and has L-time or 2-time precisely when k is on while $k+1$ is off, so that, for $k = 1, \dots, K-1$,

$$\tau_k^{0,on} + \tau_k^{L,on} = \tau_{k+1}^{L,off} + \tau_{k+1}^{2,off}. \quad (10)$$

Finally, base station k is on and has R-time or 2-time and base station $k+1$ is on and has L-time or 2-time exactly when both k and $k+1$ are on, which for $k = 1, \dots, K-1$ yields the identity

$$\tau_k^{R,on} + \tau_k^{2,on} = \tau_{k+1}^{L,on} + \tau_{k+1}^{2,on}. \quad (11)$$

The derivation of (7)-(11) shows that these are necessary conditions for the time allocation vector τ to be feasible. Note that these feasibility constraints depend only on the aggregate on-times $\tau_k^{0,on}$, $\tau_k^{L,on}$, $\tau_k^{R,on}$, $\tau_k^{2,on}$, and not on how they are shared between users in cell k . We now show that the above constraints are in fact also sufficient for feasibility.

Proposition 4.1

The conditions (7)-(11) are necessary and sufficient for the time allocation vector τ to be feasible.

Proof

See Appendix A. □

Now suppose we wish to determine the optimal schedule for some specific objective function defined in terms of the throughputs of the users. Relation (6) describes the (normalized) throughput of each user as a function of the time allocations τ . The above proposition establishes necessary and sufficient conditions for the time allocations to be feasible. We can thus formulate a mathematical programming problem governing the optimal schedule, with the time allocations playing the role of decision variables. The mathematical program may be augmented with additional constraints, e.g. further restrictions on some of the time allocations, or lower and upper bounds on the throughputs of individual users.

Observe that the throughput of each user is a linear function of the time allocation variables. The feasibility constraints (7)-(11) are also linear in the time allocation variables. Thus if the objective function is also linear, e.g., a weighted combination of the throughputs of the users, then the resulting programming problem takes the form of a linear program.

In the present paper, however, we are primarily interested in gaining qualitative insights, rather than numerically determining the optimal schedule for a specific objective function or particular parameter values. We only assume that the objective function is increasing in the throughput of each user, i.e. more throughput is always better. Henceforth, we use τ^* to denote an optimal schedule with respect to an objective function that is increasing in the throughput of each user. For optimality, a base station should always be on during 0-time, since such transmissions yield additional throughput without affecting users in neighboring cells. Thus any optimal schedule must satisfy

$$\tau_k^{0,off*} = 0 \quad \text{for all } k. \quad (12)$$

We now derive some further optimality properties. We assume that the path loss coefficients β_{ik}^L , β_{ik}^R satisfy the following monotonicity properties.

Assumption 4.1

- (i) If $i < j$, then $\beta_{ik}^R < \beta_{jk}^R$ and $\beta_{ik}^L > \beta_{jk}^L$.
- (ii) If user i is closer to base station k than user j , then $\beta_{ik}^L + \beta_{ik}^R < \beta_{jk}^L + \beta_{jk}^R$.

Property (i) holds for any monotone path loss function, provided no two user positions coincide. Property (ii) may be shown to hold for a power law path loss function.

Proposition 4.2

Suppose Assumption 4.1 is satisfied. Given an on-time allocation $(\tau_k^{0,on}, \tau_k^{L,on}, \tau_k^{R,on}, \tau_k^{2,on})$, let τ^* be the optimal schedule amongst schedules that have the same time allocation. Then for any pair of users i, j in cell k ,

- (i) if $i < j$, then $\tau_{ik}^{L*} > 0$ implies $\tau_{jk}^{0*} = \tau_{jk}^{R*} = \tau_{jk}^{2*} = 0$, and similarly $\tau_{jk}^{R*} > 0$ implies $\tau_{ik}^{0*} = \tau_{ik}^{L*} = \tau_{ik}^{2*} = 0$;
- (ii) if user i is closer to the base station than user j , then $\tau_{ik}^{0*} > 0$ implies $\tau_{jk}^{2*} = 0$.

Proof

The proof uses exchange arguments that rely on the monotonicity properties of the path loss function stated in Assumption 4.1. For the details we refer to Appendix B. □

Part (i) of the above proposition implies that users on the left edge of the cell are served when the right interfering base station is on. Similarly, users in the right sector of the cell are served when the left interfering base station is on. Thus outer users on either side of the cell are served when only the neighboring base station on the opposite side is on. Part (ii) implies that inner users in the center of the cell are served when both neighbors are on. Finally, there is an intermediate set of users on either side which receive transmissions when both neighbors are off. Thus, within each cell, the optimal transmission schedule may be characterized by the leftmost user with non-zero L-time, the rightmost user with non-zero R-time, and the inner-most users on either side of the base stations that have non-zero 0-time.

The above optimality properties may be considered local in the sense that they do not require any coordination among base stations (as long as they alternate between on periods, transmitting at maximum power, and off periods). This opens up possibilities for decentralized optimization approaches. We now describe some optimality properties that do require coordination among base stations.

Proposition 4.3

Suppose Assumption 4.1 is satisfied.

- (i) For any user i in cell k , $\tau_{ik}^{L*} > 0$ implies $\tau_{ik}^{R*} = 0$.
- (ii) For any base station k , $\tau_k^{L,off*} > 0$ implies $\tau_k^{R,off*} = \tau_k^{R,on*} = \tau_k^{2,on*} = 0$, and similarly $\tau_k^{R,off*} > 0$ implies $\tau_k^{L,off*} = \tau_k^{L,on*} = \tau_k^{2,on*} = 0$.

Proof

The proof uses exchange arguments that use the feasibility constraints stated in Proposition 4.1, and can be found in Appendix C. □

5 Symmetric networks

We now consider a symmetric linear network with identical user populations in each cell. Our objective is to find the optimal symmetric schedule, which we define to be the optimal schedule amongst those that provide the same time allocations to corresponding users within each cell, i.e.,

$$\tau_{ik}^0 = \tau_i^0, \quad \tau_{ik}^L = \tau_i^L, \quad \tau_{ik}^R = \tau_i^R, \quad \tau_{ik}^2 = \tau_i^2$$

for all base stations k . The above clearly implies that

$$\tau_k^{0,on} = \tau^{0,on}, \quad \tau_k^{L,on} = \tau^{L,on}, \quad \tau_k^{R,on} = \tau^{R,on}, \quad \tau_k^{2,on} = \tau^{2,on}.$$

By (12) we know that for optimality we must have $\tau_k^{0,off} = 0$ for every k . Substituting this into the feasibility constraints (8)-(11), we obtain

$$\begin{aligned} \tau_k^{L,off} &= \tau_{k+1}^{R,off}, \\ \tau_k^{R,off} + \tau_k^{2,off} &= \tau^{0,on} + \tau^{R,on}, \\ \tau^{0,on} + \tau^{L,on} &= \tau_{k+1}^{L,off} + \tau_{k+1}^{2,off}, \\ \tau^{R,on} &= \tau^{L,on}, \end{aligned}$$

which implies

$$\tau_k^{L,off} = \tau_k^{R,off}$$

for all base stations k . Applying Proposition 4.3, we then find that $\tau_k^{L,off} = \tau_k^{R,off} = 0$, so that

$$\tau_k^{2,off} = \tau^{0,on} + \tau^{L,on}, \quad \tau^{R,on} = \tau^{L,on},$$

while from (7) we obtain

$$\tau^{0,on} + \tau^{L,on} + \tau^{R,on} + \tau^{2,on} + \tau_k^{2,off} = 1,$$

which may be rewritten as

$$2\tau^{0,on} + 3\tau^{L,on} + \tau^{2,on} = 1.$$

The feasibility constraints for optimal symmetric schedules derived above may be expressed as

$$\sum_{i=1}^M \tau_i^0 = \tau^0, \quad \sum_{i=1}^M \tau_i^L = \tau^1, \quad \sum_{i=1}^M \tau_i^R = \tau^1, \quad \sum_{i=1}^M \tau_i^2 = \tau^2, \quad (13)$$

$$2\tau^0 + 3\tau^1 + \tau^2 = 1. \quad (14)$$

We derive a further optimality property valid for symmetric networks.

Proposition 5.1

Suppose Assumption 4.1 is satisfied. Let τ^ be an optimal symmetric schedule. If $\beta_i^L + \beta_i^R < 1$, then $\tau_i^{0*} = 0$, and if $\beta_i^L + \beta_i^R > 1$, then $\tau_i^{2*} = 0$.*

Proof

The proof, which can be found in Appendix D, uses exchange arguments that rely on the feasibility constraints stated above. □

We now describe a simple search algorithm for finding the optimal symmetric schedule with respect to a particular objective function. Suppose we wish to maximize throughput subject to certain target throughput ratios ρ_1, \dots, ρ_M , i.e.,

$$\begin{aligned} \max \quad & T \\ \text{sub} \quad & \tau_i^0 + \frac{\tau_i^L}{1 + \beta_i^L} + \frac{\tau_i^R}{1 + \beta_i^R} + \frac{\tau_i^2}{1 + \beta_i^L + \beta_i^R} = \rho_i T, \quad i = 1, \dots, M \\ & 2\tau^0 + 3\tau^1 + \tau^2 = 1, \end{aligned}$$

with the time allocation variables satisfying the identity relations (13). Here the (ρ_1, \dots, ρ_M) vector can be seen as representing the relative priorities given to the throughputs of various users, thus reflecting a certain fairness criterion. The above optimization problem may be equivalently formulated as minimizing the ‘time budget’ required to achieve the target throughput vector, i.e.,

$$\begin{aligned} \min \quad & 2\tau^0 + 3\tau^1 + \tau^2 \\ \text{sub} \quad & \tau_i^0 + \frac{\tau_i^L}{1 + \beta_i^L} + \frac{\tau_i^R}{1 + \beta_i^R} + \frac{\tau_i^2}{1 + \beta_i^L + \beta_i^R} = \rho_i, \quad i = 1, \dots, M, \end{aligned}$$

where τ_i ’s satisfy (13).

Proposition 4.2 implies that in the optimal schedule there is a user l^* such that all the users $j > l^*$ to the right of it have only non-zero L-time, so that $\tau_j^{L*} = \rho_j(1 + \beta_j^L)$, while all the users $j < l^*$ to the left of it have zero L-time. Thus l^* may be related to τ^{1*} via the relation

$$l^* = \min \left\{ i : \sum_{j=i+1}^M \rho_j(1 + \beta_j^L) \leq \tau^{1*} \right\},$$

and the amount of throughput received by l^* in L-time is

$$\hat{\rho}_{l^*} = \frac{1}{1 + \beta_{l^*}^L} \left(\tau^{1*} - \sum_{j=l^*+1}^M \rho_j (1 + \beta_j^L) \right).$$

Similarly, there is a user r^* such that all the users $j < r^*$ to the left of it have only non-zero R-time, so that $\tau_j^{R*} = \rho_j (1 + \beta_j^R)$, while all the users $j > r^*$ to the right of it have zero R-time. As a consequence

$$r^* = \min \left\{ i : \sum_{j=1}^{i-1} \rho_j (1 + \beta_j^R) \leq \tau^{1*} \right\},$$

and the amount of throughput provided to r^* in R-time is

$$\hat{\rho}_{r^*} = \frac{1}{1 + \beta_{r^*}^R} \left(\tau^{1*} - \sum_{j=1}^{r^*-1} \rho_j (1 + \beta_j^R) \right).$$

Proposition 5.1 implies that users j with $\beta_j^L + \beta_j^R < 1$ have zero 0-time. Also, users j with $\beta_j^L + \beta_j^R > 1$ have zero 2-time. The value of the objective function may thus be expressed in terms of τ^{1*} as follows

$$\begin{aligned} F(\tau^{1*}) &= 3\tau^{1*} + \sum_{j=r^*+1}^{l^*-1} \rho_j \min\{2, 1 + \beta_j^L + \beta_j^R\} \\ &+ \left[\rho_{l^*} - \frac{1}{1 + \beta_{l^*}^L} \left(\tau^{1*} - \sum_{j=l^*+1}^M \rho_j (1 + \beta_j^L) \right) \right] \min\{2, 1 + \beta_{l^*}^L + \beta_{l^*}^R\} \\ &+ \left[\rho_{r^*} - \frac{1}{1 + \beta_{r^*}^R} \left(\tau^{1*} - \sum_{j=1}^{r^*-1} \rho_j (1 + \beta_j^R) \right) \right] \min\{2, 1 + \beta_{r^*}^L + \beta_{r^*}^R\}. \end{aligned} \quad (15)$$

Thus we have reduced the original optimization problem to a minimization problem in a single variable. The function $F(\cdot)$ is linear as long as the identities of the two users l^* and r^* do not change. In other words, the function $F(\cdot)$ is piecewise linear, each segment corresponding to a pair of users l^* and r^* . At each of the transition points, either l^* decreases by 1, or r^* increases by 1, so there are exactly M segments since $r^* \leq l^*$. Thus there are $M+1$ potentially optimal schedules, corresponding to each of the $M-1$ transition points, plus the two extreme points. In principle, we could simply evaluate the objective function for each of these candidate schedules, and select the best one. However, if we could prove that the function $F(\cdot)$ is convex, then we could conclude that the m th schedule is optimal, once we have established that the $(m+1)$ th schedule does not give any improvement, so that we do not need to evaluate the remaining candidate schedules. From (15) the slope of the m th segment is

$$G(m) = 3 - \frac{1}{1 + \beta_{l_m}^L} \min\{2, 1 + \beta_{l_m}^L + \beta_{l_m}^R\} - \frac{1}{1 + \beta_{r_m}^R} \min\{2, 1 + \beta_{r_m}^L + \beta_{r_m}^R\},$$

with l_m and r_m indexing the two users l^* and r^* corresponding to the m th segment. To show that the function $F(\cdot)$ is convex, we need to prove that the slope of the segments is increasing in m .

Lemma 5.1

The function $G(m)$ is increasing in m .

Proof

Notice that $r_m \leq r_{m+1} \leq l_{m+1} \leq l_m$, so that from Assumption 4.1, we have

$$\beta_{r_m}^L \geq \beta_{r_{m+1}}^L \geq \beta_{l_{m+1}}^L \geq \beta_{l_m}^L \quad \text{and} \quad \beta_{r_m}^R \leq \beta_{r_{m+1}}^R \leq \beta_{l_{m+1}}^R \leq \beta_{l_m}^R.$$

Thus

$$\begin{aligned} G(m) &= 3 - \frac{1}{1 + \beta_{l_m}^L} \min \left\{ 2, 1 + \beta_{l_m}^L + \beta_{l_m}^R \right\} - \frac{1}{1 + \beta_{r_m}^R} \min \left\{ 2, 1 + \beta_{r_m}^L + \beta_{r_m}^R \right\} \\ &= 3 - \min \left\{ \frac{2}{1 + \beta_{l_m}^L}, 1 + \frac{\beta_{l_m}^R}{1 + \beta_{l_m}^L} \right\} - \min \left\{ \frac{2}{1 + \beta_{r_m}^R}, 1 + \frac{\beta_{r_m}^L}{1 + \beta_{r_m}^R} \right\} \\ &\leq 3 - \min \left\{ \frac{2}{1 + \beta_{l_{m+1}}^L}, 1 + \frac{\beta_{l_{m+1}}^R}{1 + \beta_{l_{m+1}}^L} \right\} - \min \left\{ \frac{2}{1 + \beta_{r_{m+1}}^R}, 1 + \frac{\beta_{r_{m+1}}^L}{1 + \beta_{r_{m+1}}^R} \right\} \\ &= 3 - \frac{1}{1 + \beta_{l_{m+1}}^L} \min \left\{ 2, 1 + \beta_{l_{m+1}}^L + \beta_{l_{m+1}}^R \right\} - \frac{1}{1 + \beta_{r_{m+1}}^R} \min \left\{ 2, 1 + \beta_{r_{m+1}}^L + \beta_{r_{m+1}}^R \right\} \\ &= G(m+1). \end{aligned}$$

□

The search algorithm may now be summarized as follows.

1. Set $m = 1$, $l_m = M$, $r_m = 1$.
 2. Compute $G(m)$.
 3. If $G(m) \geq 0$, then the m th candidate schedule is optimal.
For $j = r_m, \dots, l_m$, if $\beta_j^L + \beta_j^R > 1$, then $\tau_j^{0*} = \rho_j$;
else $\tau_j^{2*} := \rho_j(1 + \beta_j^L + \beta_j^R)$.
 4. If $G(m) < 0$, then the m th candidate schedule is not optimal.
Set $\delta^L = \rho_{l_m}(1 + \beta_{l_m}^L)$, $\delta^R = \rho_{r_m}(1 + \beta_{r_m}^R)$, $\delta = \min\{\delta^L, \delta^R\}$.
If $\delta^L < \delta^R$, then $l_{m+1} = l_m - 1$, $r_{m+1} = r_m$, $\rho_{r_m} = \rho_{r_m} - \delta(1 + \beta_{r_m}^R)$;
else $l_{m+1} = l_m$, $r_{m+1} = r_m + 1$, $\rho_{l_m} = \rho_{l_m} - \delta(1 + \beta_{l_m}^L)$.
Set $\tau_{l_m}^{L*} = \tau_{l_m}^{L*} + \delta$, $\tau_{r_m}^{R*} = \tau_{r_m}^{R*} + \delta$, $m = m + 1$.
- Return to step 2.

6 Numerical results

We now describe some numerical experiments which were conducted to quantify the performance gains from intra-cell and inter-cell scheduling. In intra-cell scheduling, the base stations are required to be on all the time, so that the optimal strategy may be determined in a distributed manner. In inter-cell scheduling, base stations are allowed to alternate between on-periods and off-periods in a coordinated manner, as outlined in the previous sections. Our main thrust in the experiments in Sections 6.1 and 6.2 is on examining the improvement in throughput obtained by these schemes over conventional CDMA. As discussed earlier, the throughputs of the intra-cell and inter-cell scheduling strategies serve as upper bounds on the achievable throughput, as they ignore constraints like the requirement of a minimum spreading factor, synchronization and pilot signals, etc. However, in Section 6.3 we also present the results of experiments in which we take some of these constraints into account. Specifically, we consider hybrid schemes that combine elements of CDMA with that of intra-cell scheduling that meet the spreading factor constraint.

Throughout we consider a linear array of cells. In the first three subsections, we focus on the case in which the user distribution in each cell is identical and the users are uniformly spaced. In Section 6.4 we consider the scenario in which users are located only on the edge of the cell in order to specifically determine the benefits of inter-cell scheduling for edge users. This experiment also takes into account the presence of pilot signals. All our results suggest that significant gains can be obtained from inter-cell and intra-cell scheduling over conventional CDMA.

6.1 Common throughput requirement

We consider the situation where all users have a common throughput target T , and examine the maximum value of T achievable with each scheme.

Figure 2 shows the maximum achievable common throughput T for intra-cell scheduling, inter-cell scheduling, and CDMA, plotted against the number of users in each cell, assuming an orthogonality factor of $f = 1$. The results are presented in terms of the throughput per unit time obtained by each user, normalized by the system bandwidth.

The figure shows that the throughputs of inter-cell and intra-cell scheduling are quite close, and that they both outperform CDMA. Furthermore, the relative improvement over CDMA increases with the number of users.

Figure 3 shows the maximum achievable common throughput T for the three schemes, normalized by the system bandwidth, as a function of the orthogonality factor f , assuming $M = 32$ users in each cell. Note that the orthogonality factor f depends on the codes used for the various users, and even if these codes are perfectly orthogonal, there is always some degree

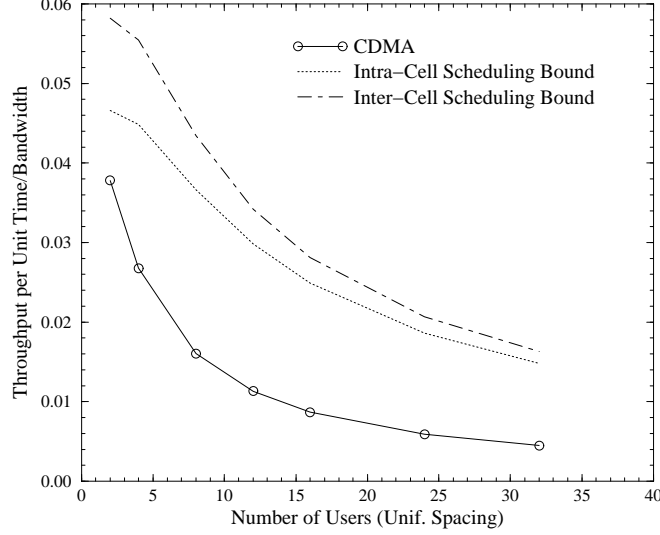


Figure 2: Maximum common throughput vs. number of users.

of interference between the signals of the various users due to multi-path. As f increases, the throughput of CDMA goes down. The throughputs of the intra-cell and inter-cell scheduling strategies are invariant with the orthogonality factor, since they involve transmission to only one user at a time. With perfect orthogonality and no multi-path ($f = 0$), the CDMA throughput is equal to that of intra-cell scheduling, since in this case the transmissions to different users do not interfere. However, the relative performance rapidly increases with f , so that values of f above 0.2 lead to an improvement of over 50% in throughput for intra-cell scheduling over CDMA.

6.2 Two-level throughput requirement

Here we consider a situation where the users closer to the base station are allowed a higher throughput than users further out. The cell is partitioned into two zones, with users in the inner zone allowed a throughput that is θ times that allowed for users in the outer zone. We refer to the latter throughput as the *basic* throughput T_b , while the throughput allowed to the users in the inner zone is referred to as the *high* throughput T_h . We call the fraction α of the cell size covered by the high throughput zone as the *high throughput coverage*.

Figure 4 plots the basic throughput T_b for the three schemes against the number of users in each cell. In this experiment we assume the orthogonality factor is $f = 1$, the high throughput coverage $\alpha = 0.75$, and the target throughput ratio $\theta = T_h/T_b = 3.0$. (Thus the users in the

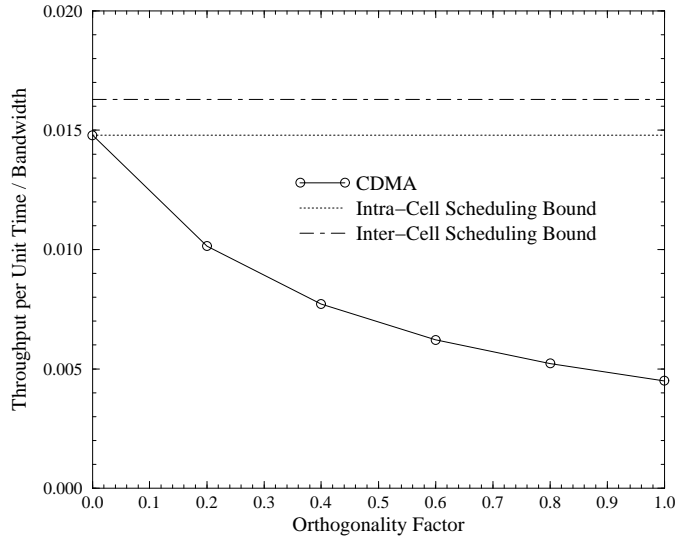


Figure 3: Maximum common throughput vs. orthogonality factor.

high throughput zone receive thrice the throughput of the users in the basic throughput zone.)

Figure 5 plots the basic throughput T_b as a function of the high throughput coverage α . In this experiment we assume there are $M = 32$ users in each cell, the orthogonality factor is $f = 1$, and the target throughput ratio $\theta = T_h/T_b = 3.0$. Notice that the basic throughput for intra-cell and inter-cell scheduling is almost constant for high throughput coverage below 0.5, and is in fact almost the same as that obtained when there is a common throughput requirement for all users (as can be seen from Figure 3). The reason for this is that the users close to the center of the cell require very little time to meet their throughput requirement, so even if their requirement is high, the time they consume is still negligible compared to that of the users near the periphery of the cell. (To some extent this remains true even with the requirement of a minimum spreading factor, which would limit the instantaneous rate of a single code.) As we increase the high throughput coverage, we allow more users further out to demand high throughput, and the achievable throughput goes down. On the other hand, the CDMA throughput goes down uniformly as the high throughput coverage is increased. Thus the scheduling schemes are more robust than CDMA, in the sense that increasing the throughput to a small number of users near the center of the cell does not affect the users that are far from the base station.

Figure 6 plots the basic throughput T_b as a function of the orthogonality factor f . In this experiment, we assume there are $M = 32$ users in each cell, the high-throughput coverage is $\alpha = 0.75$, and the target throughput ratio is $\theta = 3.0$. In the limit of perfect orthogonality

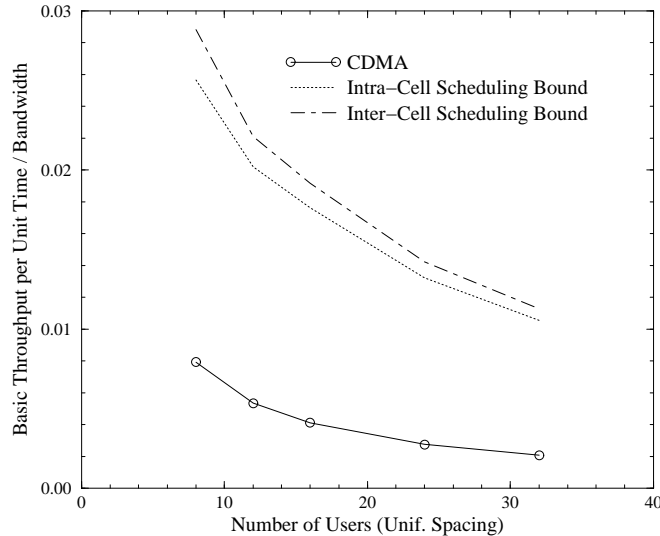


Figure 4: Basic throughput vs. number of users.

and no multi-path, the throughput performance of CDMA is identical to that of intra-cell scheduling, and drops off with increased interference between the transmissions of different users.

Figure 7 plots the basic throughput T_b as a function of the target throughput ratio θ . In this experiment we assume there are $M = 32$ users in each cell, the orthogonality factor is $f = 1$, and the high-throughput coverage $\alpha = 0.75$. The common throughput requirement scenario corresponds to a target ratio of $\theta = 1.0$. The drop in the basic throughputs of all the schemes is sharpest at low values of this ratio. Note that if the high throughput coverage were lower than 0.5, the scheduling schemes would be much less sensitive to an increase in the ratio, due to the same reason that they are insensitive to an increase in the high throughput coverage for low values of the high throughput coverage.

6.3 A hybrid CDMA/scheduling scheme

The previous results indicate a distinct advantage in throughput performance for the intra-cell and inter-cell scheduling strategies over conventional CDMA, as well as some degree of robustness. However, these results ignore the requirement of a minimum spreading factor, and indeed the bandwidth constraint itself. We now describe a simple hybrid scheme that combines the ideas of intra-cell scheduling with CDMA. This scheme does not violate these constraints, and actually approaches the performance of pure intra-cell scheduling.

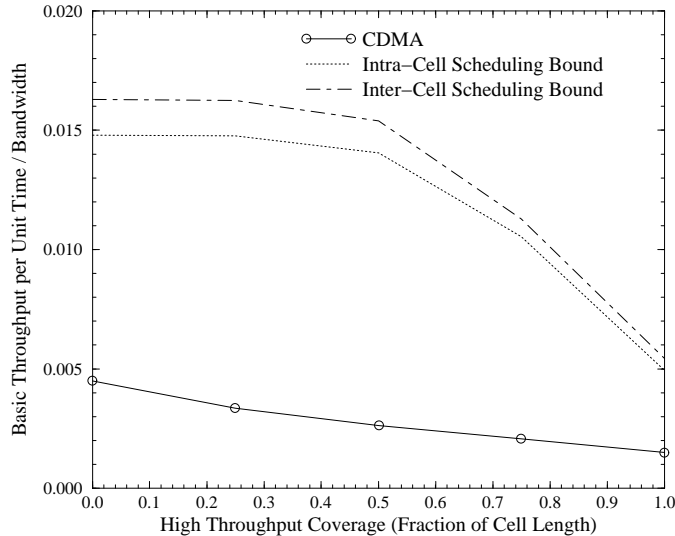


Figure 5: Basic throughput vs. high-throughput coverage α .

As noted earlier, the constraint on the minimum spreading factor imposes a limit on the maximum rate R_{\max} of a single code. When the base station transmits to the users close to it one at a time, the instantaneous rate as given by (3) may exceed R_{\max} . To prevent this, the base station could transmit to a small set of such users simultaneously, so that they receive transmissions against each other's interference, and this drives their instantaneous rate down. The size of this set has to be chosen to be large enough so that this instantaneous rate is below R_{\max} for all the users in this set. The rates that these users receive will still be high enough that the base station can meet their throughput requirements in a small amount of time. The rest of the users (those whose instantaneous rate when receiving transmission in isolation does not exceed R_{\max}) are still scheduled one at a time after the transmission to the first set of users is completed. This gives rise to a hybrid CDMA/scheduling scheme, in which the base station transmits to the first set of users in CDMA mode for part of the time, and transmits to the rest of the users in time-sharing mode during the remaining time. The time allocated to the CDMA users then depends on which set of users is chosen to be the CDMA set, and their minimum rate requirement. The orthogonality factor determines the interference between users that receive transmission simultaneously, and hence governs the minimum size of the CDMA set. In this simple strategy, each base station is on all the time, but one could also construct inter-cell scheduling strategies using this idea in which the base stations' duty cycles would be adjustable, and base stations would have the flexibility to choose which user to transmit to depending on the on/off state of the interfering base stations.

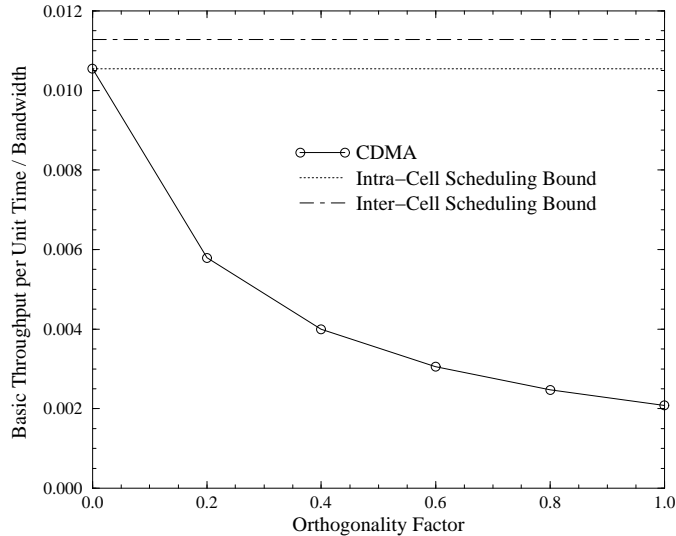


Figure 6: Basic throughput vs. orthogonality factor.

In the following experiments, we assume there are $M = 32$ users in each cell, all with a common minimum-throughput requirement T . The CDMA set consists of the N inner-most users, and we will denote by ρ the fraction $N/32$ of the total number of users that are in the CDMA set. Thus $\rho = 1$ corresponds to conventional CDMA, and $\rho = 0$ corresponds to intra-cell scheduling. The orthogonality factor is assumed to be $f = 1$, and $R_{\max} = 0.1W$ in accordance with FCC standards.

Figure 8 plots the maximum achievable common throughput T achievable with this scheme against the fraction ρ of the total number of users that are in the CDMA set. The intra-cell scheduling bound ignoring the R_{\max} constraint is also indicated. Note that the leftmost point of each curve corresponds to the minimum size that the CDMA set must have for that value of the orthogonality factor, in order that the R_{\max} constraint is met. It is seen that with the size of the CDMA set chosen to be this minimum value, this hybrid CDMA/scheduling scheme approaches within 15% of the intra-cell scheduling bound, and is significantly better than the pure CDMA strategy (which corresponds to $\rho = 1$). This shows that the intra-cell scheduling bound is a meaningful bound and can be approached quite closely even by simple schemes like the one outlined above.

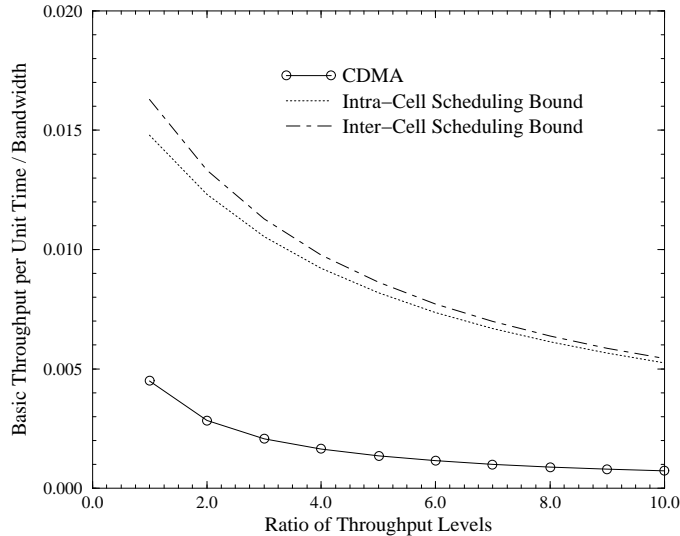


Figure 7: Basic throughput vs. target throughput ratio θ .

6.4 Benefits of inter-cell scheduling to edge users

In this section, we study the potential gains which users near the edge of a cell may derive from inter-cell scheduling. To obtain results which can be interpreted in a straightforward fashion, we suppose that there are only edge users on the two boundaries of each cell, as shown in Figure 9. (Here cell 0 is using L-time while cell -1 is using R-time. In the next period cell -1 will switch off, cell 0 will use R-time, and cell 1 will use L-time.)

To eliminate the benefits of *intra-cell* scheduling, we suppose that there is perfect orthogonality between users within a cell so that scheduling within a cell cannot increase throughput. Our results show that if all the users have a common throughput requirement, then the optimal scheduling scheme would have each base station active 2/3 of the time and idle for the remaining period. Moreover each base station's active period would be 1/3 of a scheduling interval ahead of the base station to the right and 1/3 of a scheduling interval behind the one to the left. Thus the active period is divided into equal periods of L-time and R-time. We suppose that there is a pilot signal which cannot be turned off and which accounts for 20% of the total transmit power.

Assuming self-noise is negligible, the maximum common throughput obtained by a user under a scheduled strategy is equal to

$$R_{sched} = \frac{W}{\gamma} \frac{0.8P}{3M(0.2P\mathcal{G} + \eta)},$$

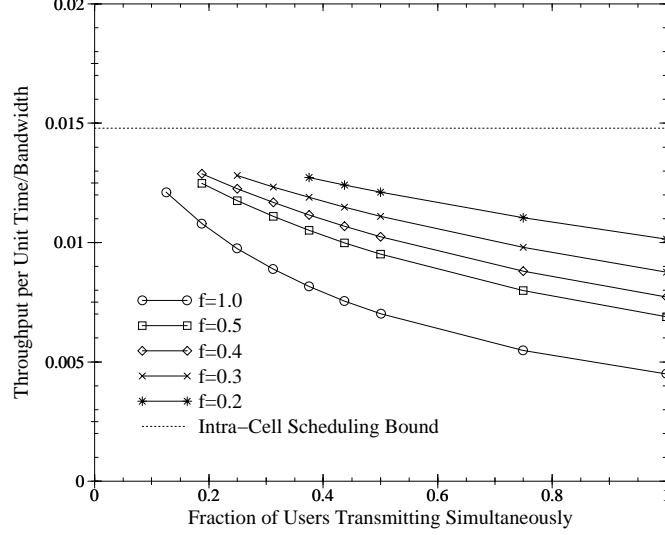


Figure 8: Maximum common throughput of hybrid CDMA/scheduling scheme vs. fraction of total number of users that are in the CDMA set.

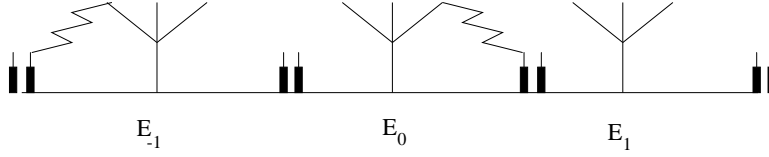


Figure 9: Inter-cell scheduling.

where M is the number of users on each edge of the cell and \mathcal{G} is the common path loss of users to the nearest interfering base station. Note that it is irrelevant whether the M users on the edge are transmitted to simultaneously or one at a time, provided the spreading constraint is not broken in the latter case.

If a pure CDMA strategy is adopted, then all the base stations are on all the time, and so the rate obtained is

$$R_{CDMA} = \frac{W}{\gamma} \frac{0.8P}{2M(P\mathcal{G} + \eta)}.$$

The ratio of throughputs obtained in each of the schemes is given by

$$R_{sched}/R_{CDMA} = \frac{2 \times (P\mathcal{G} + \eta)}{3 \times (0.2P\mathcal{G} + \eta)}.$$

Neglecting the background noise in comparison with the interference due to pilot, the ratio is approximately 10/3.

7 Conclusion

We identified optimality properties for scheduling downlink transmissions to data users in CDMA networks. We showed that under certain assumptions for arbitrary-topology networks it is optimal for a base station to transmit to only one data user at a time. Moreover, for data-only networks, we proved that in the absence of self-noise a base station, when on, should transmit at maximum power for optimality. We used these two properties to obtain a mathematical programming formulation for determining the optimal transmission schedule in linear data-only networks, with time allocations playing the role of decision variables. The optimality conditions imply that there exist (i) subsets of outer users on either side of the cell that should be served when only the neighboring base station on the opposite side is on; (ii) a subset of inner users in the center of the cell that should be served when both neighbors are on; (iii) a subset of users in the intermediate regions that should receive transmissions when both neighbors are off. Exploiting the structural properties, we derived a simple search algorithm for finding the optimal transmission schedule in symmetric scenarios. Numerical experiments illustrated that scheduling achieves substantial capacity gains over conventional CDMA. Voice users were excluded from most of the analysis. Extension of the results to a mixed voice and data system is the topic of current investigations.

As noted above, our results rely on certain idealizing assumptions that were discussed in Section 2. However, the qualitative properties of the optimal schedules are pertinent in less ideal circumstances as well. Although the actual gains from scheduling may reduce when practical constraints are taken into consideration, our numerical results show that the benefits are still substantial. For example, we considered a hybrid intra-cell scheduling scheme in Section 6.3, in which only the users on the edge of the cell are scheduled. The users in the center of the cell are transmitted to simultaneously, so that the maximum rate constraint is satisfied. Numerical results showed that the throughput closely approaches the intra-cell scheduling bound which ignored the minimum spreading factor requirement. This demonstrates that the insights from the analysis can be used in devising simple, practical schemes which are nearly optimal.

Numerical results also showed that the intra-cell scheduling bound closely approaches the inter-cell scheduling bound. This may suggest that the additional performance gains from inter-cell scheduling may not warrant the implementation complexity and the overhead introduced by the need for coordination among the base stations. On the other hand, the numerical results in Section 6.4 demonstrate that inter-cell scheduling becomes important when edge users must be provided reasonable throughputs.

In general, we observed that the performance gains from scheduling strongly depend on the degree of multi-path and self-noise. If there is little multi-path, and the base station uses orthogonal codes to transmit to distinct users, then intra-cell scheduling is indistinguishable from

conventional CDMA. (This is in contrast to the uplink, where transmission is asynchronous and orthogonality cannot be achieved, so that the case for scheduling is clearer. However, as mentioned in the introduction, the uplink is less likely to be critical for data traffic since it is commonly expected that the bulk of the traffic will be directed from the base station to the user.) When there is multi-path, scheduling may help considerably improve performance, provided the self-noise is relatively small. If the self-noise is large, intra-cell scheduling is close in performance to conventional CDMA.

Finally, we mention that scheduling has several other implications which deserve further research. For example, scheduling increases the burstiness of transmissions because of the constant rate changes. This burstiness may have a significant impact on coding techniques, power control algorithms, and end-to-end flow control protocols such as TCP. These and other issues need to be resolved before the benefits from scheduling can be fully assessed.

References

- [1] M. Abdulrahman, D. D. Falconer, A. U. H. Sheikh (1992). Equalization for interference cancellation in spread spectrum multiple access systems. In: *Proc. 42nd IEEE VTC*, 71–74.
- [2] T. Alanko, M. Kojo, H. Laamanen, M. Liljeberg, M. Moilanen, K. Raatikainen (1994). Measured performance of data transmission over cellular telephone networks. *Comp. Commun. Rev. (USA)* **24**, 24–44.
- [3] D. M. Andrews, S. C. Borst, F. Dominique, P. R. Jelenković, K. Kumaran, K. G. Ramakrishnan, P. A. Whiting (1998). Dynamic bandwidth allocation algorithms for high-speed data wireless networks. *Bell Labs Technical Journal* **3**, 30–49.
- [4] L. J. Cimini, J. C.-I. Chuang, Jr., N. R. Sollenberger (1998). Advanced cellular Internet service (ACIS). *IEEE Communications Magazine* **36**, 150–159.
- [5] J. M. Holtzman, S. Ramakrishna (1998). A scheme for throughput maximization in a dual-class CDMA system. *IEEE J. Sel. Areas Commun.* **40**, 830–844.
- [6] C.-L. I, S. Nanda (1996). Load and interference based demand assignment for wireless CDMA networks. In: *Proc. INFOCOM '96*, 235–241.
- [7] D. N. Knisely, S. Kumar, S. Laha, S. Nanda (1998). Evolution of wireless data services: IS-95 to cdma2000. *IEEE Communications Magazine* **36**, 140–149.
- [8] U. Madhow, M. L. Honig (1994). MMSE interference suppression for direct-sequence spread-spectrum CDMA. *IEEE Trans. Commun.* **42**, 3178–3188.

- [9] C. J. Mathias (1997). Wireless data – What’s real? What’s wrong? *Business Commun. Rev. (USA)* **27**, 52–54.
- [10] G. J. Pottie (1995). System design choices in personal communications. *IEEE Pers. Commun.* **2**, 50–67.
- [11] J. G. Proakis (1989). *Digital Communications*. McGraw-Hill, New York.
- [12] S. Ramakrishna (1998). *Optimal Scheduling of CDMA Systems*. PhD thesis, WINLAB, Rutgers University.
- [13] A. J. Viterbi (1995). *Principles of Spread Spectrum Communication*. Addison Wesley, Reading, Massachusetts.

A Proof of Proposition 4.1

Proposition 4.1

The conditions (7)-(11) are necessary and sufficient for the time allocation vector τ to be feasible.

Proof

The derivation of (7)-(11) shows that these are necessary conditions for feasibility. We now show they are also sufficient. Define $\tau_k^{\epsilon_{k-1}, \epsilon_k, \epsilon_{k+1}}$ as the fraction of time that the on/off state of base stations $(k-1, k, k+1)$ is $(\epsilon_{k-1}, \epsilon_k, \epsilon_{k+1}) \in \{0, 1\}^3$, $k = 1, \dots, K$. These time fractions are uniquely determined by the time allocations,

$$\begin{aligned} \tau_k^{0,0,0} &= \tau_k^{0,off}, & \tau_k^{0,0,1} &= \tau_k^{R,off}, & \tau_k^{0,1,0} &= \tau_k^{0,on}, & \tau_k^{1,0,0} &= \tau_k^{L,off}, \\ \tau_k^{0,1,1} &= \tau_k^{R,on}, & \tau_k^{1,0,1} &= \tau_k^{2,off}, & \tau_k^{1,1,0} &= \tau_k^{L,on}, & \tau_k^{1,1,1} &= \tau_k^{2,on}. \end{aligned}$$

The on/off configuration of the entire network may be represented by a vector $\delta = (\delta_0, \delta_1, \dots, \delta_K, \delta_{K+1}) \in \{0, 1\}^{K+2}$, with δ_k indicating the on/off state of base station k . For any $\delta \in \{0, 1\}^{K+2}$, let α_δ be the fraction of time that the on/off configuration of the network is δ . Then we need to show that there are $\alpha_\delta \geq 0$, with $\sum_{\delta \in \{0,1\}^{K+2}} \alpha_\delta = 1$, such that

$$\sum_{\delta \in \{0,1\}^{K+2}: (\delta_{m-1}, \delta_m, \delta_{m+1}) = (\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1})} \alpha_\delta = \tau_m^{\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}}, \quad (16)$$

for all $m = 1, \dots, K$. Define $\tau_{k,k+1}^{\epsilon_k, \epsilon_{k+1}}$ as the fraction of time that the on/off state of base stations $(k, k+1)$ is $(\epsilon_k, \epsilon_{k+1})$. In view of (8)-(11), we have

$$\tau_{k,k+1}^{\epsilon_k, \epsilon_{k+1}} = \sum_{\epsilon_{k-1}=0}^1 \tau_k^{\epsilon_{k-1}, \epsilon_k, \epsilon_{k+1}} = \sum_{\epsilon_{k+2}=0}^1 \tau_k^{\epsilon_k, \epsilon_{k+1}, \epsilon_{k+2}}.$$

Now suppose we set

$$\alpha_\delta = \frac{\prod_{k=1}^K \tau_k^{\delta_{k-1}, \delta_k, \delta_{k+1}}}{\prod_{k=1}^{K-1} \tau_{k,k+1}^{\delta_k, \delta_{k+1}}}. \quad (17)$$

Substituting (17) into (16), we obtain

$$\begin{aligned} & \sum_{\delta \in \{0,1\}^{K+2}: (\delta_{m-1}, \delta_m, \delta_{m+1}) = (\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1})} \alpha_\delta = \sum_{\delta: (\delta_{m-1}, \delta_m, \delta_{m+1}) = (\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1})} \frac{\prod_{k=1}^K \tau_k^{\delta_{k-1}, \delta_k, \delta_{k+1}}}{\prod_{k=1}^{K-1} \tau_{k,k+1}^{\delta_k, \delta_{k+1}}} \\ &= \sum_{\delta: (\delta_{m-1}, \delta_m, \delta_{m+1}) = (\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1})} \prod_{k=1}^{m-1} \frac{\tau_k^{\delta_{k-1}, \delta_k, \delta_{k+1}}}{\tau_{k,k+1}^{\delta_k, \delta_{k+1}}} \tau_m^{\delta_{m-1}, \delta_m, \delta_{m+1}} \prod_{m+1}^K \frac{\tau_k^{\delta_{k-1}, \delta_k, \delta_{k+1}}}{\tau_{k-1,k}^{\delta_{k-1}, \delta_k}} \\ &= \sum_{\delta_0=0}^1 \dots \sum_{\delta_{m-2}=0}^1 \sum_{\delta_{m+2}=0}^1 \dots \sum_{\delta_{K+1}=0}^1 \frac{\tau_1^{\delta_0, \delta_1, \delta_2}}{\tau_{1,2}^{\delta_1, \delta_2}} \dots \frac{\tau_{m-3}^{\delta_{m-4}, \delta_{m-3}, \delta_{m-2}}}{\tau_{m-3, m-2}^{\delta_{m-3}, \delta_{m-2}}} \frac{\tau_{m-2}^{\delta_{m-3}, \delta_{m-2}, \epsilon_{m-1}}}{\tau_{m-2, m-1}^{\delta_{m-2}, \epsilon_{m-1}}} \frac{\tau_{m-1}^{\delta_{m-2}, \epsilon_{m-1}, \epsilon_m}}{\tau_{m-1, m}^{\epsilon_{m-1}, \epsilon_m}} \\ & \quad \tau_m^{\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}} \frac{\tau_{m+1}^{\epsilon_m, \epsilon_{m+1}, \delta_{m+2}}}{\tau_{m, m+1}^{\epsilon_m, \epsilon_{m+1}}} \frac{\tau_{m+2}^{\epsilon_{m+1}, \delta_{m+2}, \delta_{m+3}}}{\tau_{m+1, m+2}^{\epsilon_{m+1}, \delta_{m+2}}} \frac{\tau_{m+3}^{\delta_{m+2}, \delta_{m+3}, \delta_{m+4}}}{\tau_{m+2, m+3}^{\delta_{m+2}, \delta_{m+3}}} \dots \frac{\tau_K^{\delta_{K-1}, \delta_K, \delta_{K+1}}}{\tau_{K-1, K}^{\delta_{K-1}, \delta_K}} \\ &= \tau_m^{\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}} \sum_{\delta_{m-2}=0}^1 \frac{\tau_{m-1}^{\delta_{m-2}, \epsilon_{m-1}, \epsilon_m}}{\tau_{m-1, m}^{\epsilon_{m-1}, \epsilon_m}} \dots \sum_{\delta_{m-3}=0}^1 \frac{\tau_{m-2}^{\delta_{m-3}, \delta_{m-2}, \epsilon_{m-1}}}{\tau_{m-2, m-1}^{\delta_{m-2}, \epsilon_{m-1}}} \sum_{\delta_{m-4}=0}^1 \frac{\tau_{m-3}^{\delta_{m-4}, \delta_{m-3}, \delta_{m-2}}}{\tau_{m-3, m-2}^{\delta_{m-3}, \delta_{m-2}}} \dots \sum_{\delta_0=0}^1 \frac{\tau_1^{\delta_0, \delta_1, \delta_2}}{\tau_{1,2}^{\delta_1, \delta_2}} \\ & \quad \sum_{\delta_{m+2}=0}^1 \frac{\tau_{m+1}^{\epsilon_m, \epsilon_{m+1}, \delta_{m+2}}}{\tau_{m, m+1}^{\epsilon_m, \epsilon_{m+1}}} \sum_{\delta_{m+3}=0}^1 \frac{\tau_{m+2}^{\epsilon_{m+1}, \delta_{m+2}, \delta_{m+3}}}{\tau_{m+1, m+2}^{\epsilon_{m+1}, \delta_{m+2}}} \sum_{\delta_{m+4}=0}^1 \frac{\tau_{m+3}^{\delta_{m+2}, \delta_{m+3}, \delta_{m+4}}}{\tau_{m+2, m+3}^{\delta_{m+2}, \delta_{m+3}}} \dots \sum_{\delta_{K+1}=0}^1 \frac{\tau_K^{\delta_{K-1}, \delta_K, \delta_{K+1}}}{\tau_{K-1, K}^{\delta_{K-1}, \delta_K}} \\ &= \tau_m^{\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}} \end{aligned}$$

for all $m = 1, \dots, K$, as required.

Summing the above relation over $(\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}) \in \{0,1\}^3$ and using (7), we also find that

$$\sum_{\delta \in \{0,1\}^{K+2}} \alpha_\delta = \sum_{\epsilon_{m-1}=0}^1 \sum_{\epsilon_m=0}^1 \sum_{\epsilon_{m+1}=0}^1 \sum_{\delta: (\delta_{m-1}, \delta_m, \delta_{m+1}) = (\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1})} \alpha_\delta = \sum_{\epsilon_{m-1}=0}^1 \sum_{\epsilon_m=0}^1 \sum_{\epsilon_{m+1}=0}^1 \tau_m^{\epsilon_{m-1}, \epsilon_m, \epsilon_{m+1}} = 1.$$

□

B Proof of Proposition 4.2

In this and the following sections of the appendix, we use $a \wedge b$ to denote the minimum of a and b .

Proposition 4.2

Suppose Assumption 4.1 is satisfied. Given an on-time allocation $(\tau_k^{0,on}, \tau_k^{L,on}, \tau_k^{R,on}, \tau_k^{2,on})$, let τ^* be the optimal schedule amongst schedules that have the same time allocation. Then for any pair of users i, j in cell k ,

- (i) if $i < j$, then $\tau_{ik}^{L*} > 0$ implies $\tau_{jk}^{0*} = \tau_{jk}^{R*} = \tau_{jk}^{2*} = 0$, and similarly $\tau_{jk}^{R*} > 0$ implies $\tau_{ik}^{0*} = \tau_{ik}^{L*} = \tau_{ik}^{2*} = 0$;
- (ii) if user i is closer to the base station than user j , then $\tau_{ik}^{0*} > 0$ implies $\tau_{jk}^{2*} = 0$.

Proof

For convenience we omit the base station index k from the notation.

Part (i). Consider a pair of users i, j with $i < j$ and $\tau_i^{L*} > 0$.

First suppose $\tau_j^{0*} > 0$. We can then achieve higher throughput by exchanging L-time and 0-time between users i and j as follows. Take some positive $\delta \leq \tau_i^{L*} \wedge \tau_j^{0*}(1 + \beta_i^L)$. Construct the time allocation vector τ' from τ^* by modifying the L-time and 0-time allotted to users i and j by setting

$$\tau_i^{L'} = \tau_i^{L*} - \delta, \quad \tau_j^{L'} = \tau_j^{L*} + \delta, \quad \tau_i^{0'} = \tau_i^{0*} + \delta \frac{1}{1 + \beta_i^L}, \quad \tau_j^{0'} = \tau_j^{0*} - \delta \frac{1}{1 + \beta_i^L},$$

but not changing any other components of τ^* . Clearly none of the users other than i and j are affected. Moreover, since allocations are just being interchanged between users i and j in the same cell, the aggregate on-times remain the same and thus the new allocation vector τ' is also feasible. The throughput of user i is not altered either, because the gain in 0-time exactly compensates for the loss in L-time. The change in the (normalized) throughput of user j is

$$\frac{\delta}{1 + \beta_j^L} - \frac{\delta}{1 + \beta_i^L} = \delta \frac{\beta_i^L - \beta_j^L}{(1 + \beta_i^L)(1 + \beta_j^L)} > 0$$

since $\beta_i^L > \beta_j^L$. Thus the throughput vector for τ' dominates that for τ^* , contradicting the optimality of τ^* .

Now suppose $\tau_j^{R*} > 0$. We can then improve throughput by swapping L-time and R-time between users i and j as follows. Choose some positive $\delta \leq \tau_i^{L*} \wedge \tau_j^{R*}(1 + \beta_i^L)/(1 + \beta_i^R)$. Construct the time allocation vector τ' from τ^* by setting

$$\tau_i^{L'} = \tau_i^{L*} - \delta, \quad \tau_j^{L'} = \tau_j^{L*} + \delta, \quad \tau_i^{R'} = \tau_i^{R*} + \delta \frac{1 + \beta_i^R}{1 + \beta_i^L}, \quad \tau_j^{R'} = \tau_j^{R*} - \delta \frac{1 + \beta_i^R}{1 + \beta_i^L}.$$

Again, none of the users other than i and j are affected. The throughput of user i does not change either, since the gain in R-time makes up for the loss in L-time. The increase in (normalized) throughput for user j is

$$\frac{\delta}{1 + \beta_j^L} - \frac{\delta}{1 + \beta_j^R} \frac{1 + \beta_i^R}{1 + \beta_i^L} = \delta \frac{(1 + \beta_i^L)(1 + \beta_j^R) - (1 + \beta_j^L)(1 + \beta_i^R)}{(1 + \beta_i^L)(1 + \beta_j^L)(1 + \beta_j^R)} > 0$$

because $\beta_i^L > \beta_j^L$ and $\beta_i^R < \beta_j^R$.

Finally suppose $\tau_j^{2*} > 0$. We can then improve throughput by transferring L-time and 2-time between users i and j as follows. Choose some positive $\delta \leq \tau_i^{L*}, \tau_j^{2*}(1 + \beta_i^L)/(1 + \beta_i^L + \beta_i^R)$. Construct the time allocation vector τ' from τ^* by setting

$$\tau_i^{L'} = \tau_i^{L*} - \delta, \quad \tau_j^{L'} = \tau_j^{L*} + \delta, \quad \tau_i^{2'} = \tau_i^{2*} + \delta \frac{1 + \beta_i^L + \beta_i^R}{1 + \beta_i^L}, \quad \tau_j^{2'} = \tau_j^{2*} - \delta \frac{1 + \beta_i^L + \beta_i^R}{1 + \beta_i^L}.$$

Again, none of the users other than i and j are affected. The throughput of user i does not change either, since the loss in L-time is offset by the gain in 2-time. The increase in the (normalized) throughput of user j is

$$\begin{aligned} \frac{\delta}{1 + \beta_j^L} - \frac{\delta}{1 + \beta_j^L + \beta_j^R} \frac{1 + \beta_i^L + \beta_i^R}{1 + \beta_i^L} &= \delta \frac{(1 + \beta_i^L)(1 + \beta_j^L + \beta_j^R) - (1 + \beta_j^L)(1 + \beta_i^L + \beta_i^R)}{(1 + \beta_i^L)(1 + \beta_j^L)(1 + \beta_j^L + \beta_j^R)} \\ &= \delta \frac{(1 + \beta_i^L)(1 + \beta_j^R) - (1 + \beta_j^L)(1 + \beta_i^R)}{(1 + \beta_i^L)(1 + \beta_j^L)(1 + \beta_j^L + \beta_j^R)} \\ &> 0 \end{aligned}$$

because $\beta_i^L > \beta_j^L$ and $\beta_i^R < \beta_j^R$.

Thus we have shown that if $\tau_i^{L*} > 0$, then we must have $\tau_j^{0*} = \tau_j^{R*} = \tau_j^{2*} = 0$ for optimality. The second implication follows by symmetry. This completes the proof of part (i).

Part (ii). Consider a pair of users i, j with user i closer to the base station than user j and $\tau_i^{0*} > 0$. Suppose $\tau_j^{2*} > 0$. We can then achieve higher throughput by exchanging 0- and 2-time between users i and j as follows. Take some positive $\delta \leq \tau_i^{0*} \wedge \tau_j^{2*} / (1 + \beta_i^L + \beta_i^R)$. Construct the time allocation vector τ' from τ^* by modifying the 0-time and 2-time allotted to users i and j as follows:

$$\tau_i^{0'} = \tau_i^{0*} - \delta, \quad \tau_j^{0'} = \tau_j^{0*} + \delta, \quad \tau_i^{2'} = \tau_i^{2*} + \delta(1 + \beta_i^L + \beta_i^R), \quad \tau_j^{2'} = \tau_j^{2*} - \delta(1 + \beta_i^L + \beta_i^R),$$

while leaving all other components of τ^* unaltered. The throughput of all users except i and j remains the same. In fact, the throughput of user i is not altered either, because the gain in 2-time exactly compensates for the loss in 0-time. The change in (normalized) throughput for user j is

$$\delta - \delta \frac{1 + \beta_i^L + \beta_i^R}{1 + \beta_j^L + \beta_j^R} = \delta \frac{\beta_j^L + \beta_j^R - \beta_i^L - \beta_i^R}{(1 + \beta_i^L + \beta_i^R)(1 + \beta_j^L + \beta_j^R)} > 0$$

since $\beta_i^L + \beta_i^R < \beta_j^L + \beta_j^R$. Thus the throughput vector for τ' dominates that for τ^* , contradicting the optimality of τ^* . This completes the proof of part (ii). \square

C Proof of Proposition 4.3

Proposition 4.3

Suppose Assumption 4.1 is satisfied.

- (i) For any user i in cell k , $\tau_{ik}^{L*} > 0$ implies $\tau_{ik}^{R*} = 0$.
- (ii) For any base station k , $\tau_k^{L,off*} > 0$ implies $\tau_k^{R,off*} = \tau_k^{R,on*} = \tau_k^{2,on*} = 0$, and similarly $\tau_k^{R,off*} > 0$ implies $\tau_k^{L,off*} = \tau_k^{L,on*} = \tau_k^{2,on*} = 0$.

Proof

Part (i). For convenience we omit the base station index k from the notation. Consider a user i with $\tau_i^{L*} > 0$. Suppose $\tau_i^{R*} > 0$. We can then achieve higher throughput by trading in L-time and R-time for 0-time and 2-time as follows. Take some positive $\delta \leq \tau_i^{L*} \wedge \tau_i^{R*}$. Construct the time allocation vector τ' from τ^* by setting

$$\tau_i^{L'} = \tau_i^{L*} - \delta, \quad \tau_i^{R'} = \tau_i^{R*} - \delta, \quad \tau_i^{0'} = \tau_i^{0*} + \delta, \quad \tau_i^{2'} = \tau_i^{2*} + \delta,$$

and leaving all other components of τ^* unchanged. It is easily verified from conditions (7)-(11) that the modified time allocation vector τ' remains feasible. Clearly none of the users other than i are affected. The increase in (normalized) throughput of user i is

$$\begin{aligned} \delta + \frac{\delta}{1 + \beta_i^L + \beta_i^R} - \frac{\delta}{1 + \beta_i^L} - \frac{\delta}{1 + \beta_i^R} &= \delta \left[1 + \frac{1}{1 + \beta_i^L + \beta_i^R} - \frac{1}{1 + \beta_i^L} - \frac{1}{1 + \beta_i^R} \right] \\ &= \frac{\beta_i^R \beta_i^L (2 + \beta_i^L + \beta_i^R)}{(1 + \beta_i^R)(1 + \beta_i^L)(1 + \beta_i^L + \beta_i^R)} \\ &> 0, \end{aligned}$$

which contradicts the optimality of τ^* . Thus we must have $\tau_i^{R*} = 0$. This establishes part (i).

Part (ii). Consider a base station k with $\tau_k^{L,off*} > 0$.

First suppose $\tau_k^{2,on*} > 0$. We can then improve throughput by exchanging L-off-time and 2-on-time for L-on-time and 2-off-time as follows. Choose some positive $\delta \leq \tau_k^{L,off*} \wedge \tau_k^{2,on*}$. The feasibility constraint (8) and property (12) imply that $\tau_{k+1}^{R,off*} = \tau_k^{L,off*} \geq \delta$. The condition (11) ensures that there exist positive δ_1, δ_2 , with $\delta_1 + \delta_2 = \delta$, such that $\tau_{k+1}^{L,on*} \geq \delta_1$ and $\tau_{k+1}^{2,on*} \geq \delta_2$. Construct the time allocation vector τ' by setting

$$\begin{aligned} \tau_k^{L,off'} &= \tau_k^{L,off*} - \delta, & \tau_k^{2,on'} &= \tau_k^{2,on*} - \delta, & \tau_k^{L,on'} &= \tau_k^{L,on*} + \delta, & \tau_k^{2,off'} &= \tau_k^{2,off*} + \delta, \\ \tau_{k+1}^{R,off'} &= \tau_{k+1}^{R,off*} - \delta, & \tau_{k+1}^{2,off'} &= \tau_{k+1}^{2,off*} + \delta, \\ \tau_{k+1}^{L,on'} &= \tau_{k+1}^{L,on*} - \delta_1, & \tau_{k+1}^{2,on'} &= \tau_{k+1}^{2,on*} - \delta_2, & \tau_{k+1}^{0,on'} &= \tau_{k+1}^{0,on*} + \delta_1, & \tau_{k+1}^{R,on'} &= \tau_{k+1}^{R,on*} + \delta_2, \end{aligned}$$

but not changing any other components of τ^* . It is readily checked from conditions (7)-(11) that the modified time allocation vector τ' remains feasible. The throughput vector may now be improved, because 0-time is always better than L-time, and L-time and R-time is always better than 2-time. This contradicts the optimality of τ^* .

Now suppose $\tau_k^{R,on*} > 0$. We can then achieve higher throughput by converting L-off-time and R-on-time into 0-on-time and 2-off-time as follows. Take some positive $\delta \leq \tau_k^{L,off*} \wedge \tau_k^{R,on*}$. Construct the time allocation vector τ' by defining

$$\begin{aligned} \tau_k^{L,off'} &= \tau_k^{L,off*} - \delta, & \tau_k^{R,on'} &= \tau_k^{R,on*} - \delta, & \tau_k^{0,on'} &= \tau_k^{0,on*} + \delta, & \tau_k^{2,off'} &= \tau_k^{2,off*} + \delta, \\ \tau_{k+1}^{R,off'} &= \tau_{k+1}^{R,off*} - \delta, & \tau_{k+1}^{2,off'} &= \tau_{k+1}^{2,off*} + \delta, \end{aligned}$$

$$\tau_{k+1}^{L,on'} = \tau_{k+1}^{L,on*} - \delta_1, \quad \tau_{k+1}^{2,on'} = \tau_{k+1}^{2,on*} - \delta_2, \quad \tau_{k+1}^{0,on'} = \tau_{k+1}^{0,on*} + \delta_1, \quad \tau_{k+1}^{R,on'} = \tau_{k+1}^{R,on*} + \delta_2,$$

with positive $\delta_1, \delta_2, \delta_1 + \delta_2 = \delta$, such that $\tau_{k+1}^{L,on*} \geq \delta_1, \tau_{k+1}^{2,on*} \geq \delta_2$. It is easily verified from conditions (7)-(11) that the modified time allocation vector τ' remains feasible. The throughput vector may again be improved, because 0-time is always better than L-time and R-time, and R-time is always better than 2-time.

Finally suppose $\tau_k^{R,off*} > 0$. We can then improve throughput by exchanging L-off-time and R-off-time for 0-on-time and 2-off-time as follows. Choose some positive $\delta \leq \tau_k^{L,off*} \wedge \tau_k^{R,off*}$. Construct the time allocation vector τ' by setting

$$\begin{aligned} \tau_k^{L,off'} &= \tau_k^{L,off*} - \delta, & \tau_k^{R,off'} &= \tau_k^{R,off*} - \delta, & \tau_k^{0,on'} &= \tau_k^{0,on*} + \delta, & \tau_k^{2,off'} &= \tau_k^{2,off*} + \delta, \\ \tau_{k-1}^{L,off'} &= \tau_{k-1}^{L,off*} - \delta, & \tau_{k-1}^{2,off'} &= \tau_{k-1}^{2,off*} + \delta, & \tau_{k+1}^{R,off'} &= \tau_{k+1}^{R,off*} - \delta, & \tau_{k+1}^{2,off'} &= \tau_{k+1}^{2,off*} + \delta. \end{aligned}$$

It is easy to check from conditions (7)-(11) that the modified time allocation vector τ' remains feasible. Again, the throughput vector may be improved, because transmitting in 0-time is always better than not transmitting at all.

Thus, we have shown that if $\tau_k^{L,off*} > 0$, then we must have $\tau_k^{R,off*} = \tau_k^{R,on*} = \tau_k^{2,on*} = 0$. In a similar fashion, it may be shown that if $\tau_k^{R,off*} > 0$, then we must have $\tau_k^{L,on*} = \tau_k^{L,off*} = \tau_k^{2,on*} = 0$. This completes the proof of part (ii). \square

D Proof of Proposition 5.1

Proposition 5.1

Suppose Assumption 4.1 is satisfied. Let τ^* be an optimal symmetric schedule. If $\beta_i^L + \beta_i^R < 1$, then $\tau_i^{0*} = 0$, and if $\beta_i^L + \beta_i^R > 1$, then $\tau_i^{2*} = 0$.

Proof

Consider a user i with $\beta_i^L + \beta_i^R < 1$. Suppose $\tau_i^{0*} > 0$. We can then achieve higher throughput by trading in 0-time for 2-time as follows. Take some positive $\delta \leq \tau_i^{0*}$. Construct the time allocation vector τ' from τ^* by modifying the time allotment to user i into $\tau_i^{0'} = \tau_i^{0*} - \delta$, $\tau_i^{2'} = \tau_i^{2*} + 2\delta$, but not changing any other components of τ^* . It is easily verified from (13)-(14) that the modified time allocation vector τ' remains feasible. Clearly none of the users other than i are affected. The increase in (normalized) throughput of user i is

$$\frac{2\delta}{1 + \beta_i^L + \beta_i^R} - \delta = \frac{\delta(1 - \beta_i^L - \beta_i^R)}{1 + \beta_i^L + \beta_i^R} > 0,$$

which contradicts the optimality of τ^* . Thus we must have $\tau_i^{0*} = 0$. The proof of the second statement is similar. \square